REMARKS

Applicant respectfully submits the above amendments to provide additional claims, correct informalities, and improve readability, all without adding new matter.

Among the informalities corrections is a correction on line 25 page 8, in which a typographical error is corrected to thereby correctly list "T_{S2}" within the phrase "the start (time T_{S2}) of the next pre-scheduled multicast". This correction is clearly seen to be correct because the sentence on lines 26-27 of page 8 states that "A time region 410 is within the time threshold (2δ) before the start of the next pre-scheduled multicast", and FIG. 4A clearly shows that, indeed, the time T_{S2} is the start of such a next pre-scheduled multicast for which the time region 410 is within 2δ before the time T_{S2}.

Another informality correction is on line 19 of page 9, in which a typographical error is corrected to thereby correctly list "T_{S1}" within the phrase "a duration d, which equals T_A minus T_{S1}". This correction is clearly seen to be correct in FIG. 4B.

Among the readability improvements are cumulative header outline numbers (e.g., "IV.C" instead of merely "C") and improved equation labels (e.g., "(eq12)" instead of merely "(12)"). Such readability improvements are especially helpful given the automatic formatting performed by USPTO-provided software for the USPTO's Electronic Filing System.

Please note that equations are present in the amendments merely in the course of showing the amending of their equation labels. The equations themselves are NOT amended.

VERSION OF AMENDMENTS WITH MARKINGS TO SHOW CHANGES MADE (relative to the <u>originally-filed paper copy</u> of the present Application.)

IN THE SPECIFICATION:

The paragraph that spanned the bottom of page 8 and the top of page 9:

FIG. 4A is a schematic timing chart that illustrates admission and transmission for a statically admitted user. FIG. 4A shows a period of time between the start (time T_{S1}) of one pre-scheduled multicast and the start (time **[T_{S1}]** $\underline{T_{S2}}$) of the next pre-scheduled multicast. A time region 410 is within the time threshold (2δ) before the start of the next pre-scheduled multicast. A user that arrives within the time region 410 will be admitted statically. A time region 420 is not within the time threshold (2δ) before the start of the next pre-scheduled multicast. A user that arrives within the time region 420 will be admitted dynamically, as is further discussed starting in the following paragraph. The statically admitted user preferably receives the entire video from the next pre-scheduled multicast, preferably on a single multicast channel and without having to change channels, e.g., without requiring further communication, for channel establishment, between the user (e.g., client) and any router of the network that couples the client to the VoD system.

The paragraph that appeared as the second-to-last paragraph on page 9 and that starts at line 13 of page 9:

FIG. 4B is a schematic timing chart that illustrates admission and preferred transmission for a dynamically admitted user. Times T_{S1} and T_{S2} in FIG. 4B are as described in connection with FIG. 4A. Upon arriving (at time T_A), the user who will be dynamically admitted begins receiving and caching the most recently-started multicast of the video via that multicast's static multicast channel. However, the dynamically admitted user has already missed the beginning of the video. The missed amount is of a duration d, which equals T_A minus **[T_D]** $\underline{T_{S1}}$.

The paragraph that appeared at lines 5-16 of page 11:

FIG. 5 is a schematic block diagram that depicts a VoD system 100a, according to a preferred embodiment of the present invention. The VoD system 100a may be considered to be a particular embodiment of the VoD system 100 of FIG. 1. The VoD system 100a is especially scalable and suitable for very large-scale deployment, for example, to serve at least one million, at least ten million, or at least twenty million video subscribers. The VoD system 100a is preferably built by modifying a conventional video network, for example, a cable TV network, by adding the features being described. Since video networks are already a well-known conventional technology, FIG. 5 and the present document **[do]** <u>does</u> not need to describe every element of a video network. Instead, FIG. 5 and the present document concentrate on the components of a VoD system that are especially relevant to the discussed embodiments of the present invention.

The paragraph that spanned the bottom of page 13 and the top of page 14:

As mentioned above in connection with FIGS. 4B and 4C, the client to be dynamically admitted has already begun to cache video from the most recently started static multicast of the video, under direction from the AC 130a. The AC 130a knows the duration (or at least an upper bound of it) that the client to be dynamically admitted has missed (e.g., duration D1 of FIG. 4C). This duration is the minimum amount, from the client's point of view, of the front of the video that the dynamically scheduled transmission (multicast) should include. The START request that is sent (620) includes an identifier (e.g., the title or a numeric code) for the video and the value of the needed duration (e.g., duration D1 of FIG. 4C). With the single sending (620) of the START request, the AC 130a begins to participate in, and may have initiated, a current dynamic admission cycle for the video. The AC 130a records locally the value of the needed duration (e.g., duration D1 of FIG. 4C) as the maximum needed duration, mentioned above. The maximum needed duration, of the AC 130a for the video, is the greatest duration of the front portion of the video that is needed by any client of the AC 130a that

[are] is participating in the present dynamic admission cycle. After sending (620) the START request (and preferably receiving confirmation of its receipt), the AC 130a enters a second state 622.

The equation that appeared at line 1 of page 22:

$$T_R = \frac{L}{N_S}$$

(**[1]** eq1)

The paragraph that appeared at lines 3-14 of page 22:

Under the preferred pre-scheduling of the multicast transmissions, the fact that a particular pre-scheduled multicast of a movie takes place and the start time of that multicast are not in response to any user arrival or arrivals that will receive that multicast. For example, even if no user is expected to watch, or no user actually watches that multicast, the multicast still takes place at the pre-scheduled time. The pre-scheduling preferably takes place substantially in advance, for example, more than about 6 hours in advance, or more than about one day in advance, or more than about one week in advance of the **[multitask]** multicast. The schedules are conveyed to each AC 130a (of FIG. 5) so that at any time, each AC 130a has **[an]** a correct schedule of all pre-scheduled multicasts that have begun or that are next to begin for any available video whose admission and transmission are to use the above discussed methodology.

The paragraph that appeared at lines 17-26 of page 23:

The admission threshold duration 2□ may be adjusted automatically. As the system becomes busier, i.e., as the arrival rate increases, a smaller percentage of users should be admitted dynamically, and the admission threshold duration 2□ is increased. Preferably, a separate admission threshold is maintained for each particular video at a value such that the average startup latency for statically admitted users is about equal to the average startup latency for dynamically admitted users. Methodology for so maintaining the threshold by automatic

adjustment is discussed further below. Similarly, the artificial minimum **[weight]** wait Wmin in the dynamic allocation may be adjusted automatically so that as the system becomes busier, the minimum **[weight]** wait Wmin is increased.

The paragraph that spanned the bottom of page 23 and the top of page 24:

Thus The SS-VoD architecture, using the admission threshold, can easily trade off average start-up delay for capacity in a continuous manner, to exhibit very graceful degradation of performance even under exceptionally high arrival rates. Furthermore, the system performance of the preferred embodiment is, to a large degree, guaranteed such that in the worst case, assuming there are enough admission controllers for all clients, the start-up delay is no worse than the start up delay of an NVoD system that corresponds to the static channels of the SS-VoD system. (In fact, if about a 50/50 allocation of static and dynamic channels is used for a video, then the worst-case start-up delay for that video can be expected to be substantially better than merely the delay of its static channels, as will be further discussed. In sharp contrast to conventional VoD systems where the system cost can increase at least linearly with the system scale, an SS-VoD system becomes more efficient as the system **[scale]** scales up and can ultimately be scaled up to serve any number of users while still keeping the startup latency short and in fact, bounded by a worst-case latency. The system cost of an SS-VoD system grows substantially less than linearly with the number of user arrivals that are to be handled.

The header line that appeared at line 3 of page 25:

IV.A. Pause-Resume

The header line that appeared at line 21 of page 25:

IV.B. Slow Motion

The header line that appeared at line 19 of page 26:

IV.C. Seeking

The header line that appeared at line 6 of page 28:

V.A. Waiting Time for Statically-Admitted Clients

The equation that appeared at line 13 of page 28:

$$W_S(\delta) = \delta$$

([2] eq2)

The header line that appeared at line 16 of page 28:

V.B. Waiting Time for Dynamically-Admitted Clients

The equation that appeared at line 18 of page 29:

$$P_S = \frac{2\delta}{T_R}$$

([3] eq3)

The equation that appeared at line 23 of page 29:

$$\lambda_D = (1 - P_S)\lambda$$

([4] eq4)

The equation that appeared at line 2 of page 30:

$$\frac{1}{\lambda_S} = W_C(\delta) + \frac{1}{\lambda_D}$$

([5] eq5)

The equation that appeared at line 10 of page 30:

$$0 < s < T_R - 2\delta$$

([6] eq6)

The equation that appeared at line 18 of page 30:

$$W_C(\delta) = \frac{E_C(N_D, u)}{N_D(1-\rho)}\left(\frac{C_A^2 + C_S^2}{2}\right)T_S$$

([7] eq7)

The equation that appeared at line 20 of page 30:

$$C_S^2 = \frac{(T_R - 2\delta)^2}{12}\left(\frac{2}{T_R - 2\delta}\right)^2 = \frac{1}{3}$$

([8] eq8)

The equation that appeared at line 23 of page 30:

$$T_S = \frac{T_R - 2\delta}{2}$$

([9] eq9)

The equation that appeared at line 3 of page 31:

$$E_C(N_D, u) = \frac{u^{N_D} / N_D!}{u^{N_D} / N_D! + (1-\rho)\sum_{k=0}^{N_D-1}\frac{u^k}{k!}}$$

([10] eq10)

The paragraph that appeared at lines 4-8 of page 31:

Since the traffic intensity depends on the average waiting time, and the traffic intensity is needed to compute the average waiting time, Equation ([7] eq7) is in fact recursively defined. Due to ([10] eq10), Equation ([7] eq7) does not appear to be analytically solvable. Therefore, we can apply numerical methods to solve for $W_C(\delta)$ to compute numerical results.

The equation that appeared at line 19 of page 31:

$$W_2(\delta) = W_C(\delta)\left(1 - \left(\frac{1 + (T_R - 2\delta)/2W_C(\delta)}{1 - e^{\frac{-(T_R-2\delta)}{W_C(\delta)}}}\right)\frac{(T_R - 2\delta)}{W_C(\delta)} e^{\frac{-(T_R-2\delta)}{W_C(\delta)}}\right)$$

(**[11]** eq11)

The equation that appeared at line 23 of page 31:

$$W_D(\delta) = \frac{W_1(\delta) + M_2(\delta)W_2(\delta)}{1 + M_2(\delta)} = \frac{W_1(\delta) + W_C(\delta)\lambda_D W_2(\delta)}{1 + W_C(\delta)\lambda_D}$$

(**[12]** eq12)

The equation appeared at line 3 of page 32:

$$M_2(\delta) = W_C(\delta)\lambda_D$$

(**[13]** eq13)

The header line that appeared at line 19 of page 32:

V.C. Admission Threshold

The equation that appeared at line 25 of page 32:

$$\delta = \min\{x \mid (W_S(x) - W_D(x)) \le \varepsilon, T_R \ge x \ge 0\}$$

(**[14]** eq14)

The header line that appeared at line 7 of page 33:

V.D. Channel Partitioning

The paragraph that appeared at lines 8-19 of page 33:

An important configuration parameter in SS-VoD is the partitioning of available channels for use as dynamic and static multicast channels. Intuitively, having too many dynamic multicast channels will increase the traffic intensity at

the dynamic multicast channels due to increases in the service time (c.f. Equations ([1] eq1) and ([9] eq9)). On the other hand, having too few dynamic multicast channels may also result in higher load at the dynamic multicast channels. We can find the optimal channel partitioning policy by enumerating all possibilities, which in this case is of O(*N*). If the patching transmission is restricted to being a unicast (as shown in FIG. 4B), then the optimal channel partition policy is arrival-rate dependent. However, we found that, if the patching transmission is a multicast (see FIG. 4C), then the optimal channel partitioning policy is relatively independent of the user arrival rate in SS-VoD.

The header line that appeared at line 7 of page 34:

VI.A. Model Validation

The header line that appeared at line 24 of page 34:

VI.B. Channel Partitioning

The equation that appeared at line 5 of page 35:

$$\frac{w(r)}{\min\{w(r), \forall r\}}$$

([15] eq15)

The header line that appeared at line 17 of page 35:

VI.C. Latency Comparisons

The header line that appeared at line 8 of page 36:

VI.D. Channel Requirement

The header line that appeared at line 23 of page 36:

VI.E. Performance at Light Loads

The equation label that appeared at line 4 of page 37:

([16] eq16)

IN THE CLAIMS:

2-41. (New)

CONCLUSION AND SIGNATURE

If the Examiner has any questions regarding the present Application or the present Amendment or the planned Electronic Filing for Pre-Grant Publication mentioned above, the Examiner is invited to call Applicant's representative at the telephone number indicated below.

Respectfully submitted,

Date: Nov. 1, 2001

Chiahua George Yu
Reg. No. 43,301
Attorney for Applicant
Tel: (408) 739-4518
Fax: (408) 739-2300

Law Offices of C. George Yu
1250 Oakmead Parkway
Suite 210
Sunnyvale, CA 94085
U.S.A.